European Commission

JRC TECHNICAL REPORTS

Artificial Intelligence at the JRC: Survey Results

Editors: Nativi, S. and Gómez Losada, A.

2019

EUR 29803 EN

Contact information [optional element]
Stefano Nativi (Ed.)
European Commission, Joint Research Centre, TP262, Via Fermi, 21027 Ispra (VA), ITALY
stefano.nativi@ec.europa.eu
Tel.: +39 0332 785 075

EU Science Hub
https://ec.europa.eu/jrc

JRC117234

EUR 29803 EN

PDF ISBN 978-92-76-08847-9 ISSN 1831-9424 doi:10.2760/54605

Luxembourg: Publications Office of the European Union, 2019.

How to cite this report: Nativi S. and Gómez Losada A. (Ed.), *Artificial Intelligence at the JRC: Survey results*, EUR 29803 EN, Publications Office of the European Union, Luxembourg, 2019, ISBN 978-92-76-08847-9, doi:10.2760/54605, JRC117234.

Table of Contents

EXECUTIVE SUMMARY

This report presents the results of a survey on Artificial Intelligence (AI) at JRC –run from the *18th of May to the 06th of June 6 2018*.

The questionnaire was completed by *108 respondents* (74% men and 26% women) from *29 different Units*. Almost 90% were JRC *Contract Agents* and *Administrators*.

Learning-driven AI at JRC

The survey results show a well-established and on-going activity by the JRC staff, in the diverse AI areas. All the proposed macro areas are presently covered by one or more activities, at JRC. However, almost half of these activities deal with *Machine Learning* and *Deep Learning* (with a clear predominance of Machine Learning). *Knowledge Representation*, *Big Data Stewardship*, and *Natural Language Processing* together cover another 30% of the JRC activities, in the AI realm.

Extraordinary potential growth of AI areas at JRC

JRC Staff expressed an overwhelming interest in AI (as average, four AI areas were indicated by each respondents). *Deep Learning* and *Machine Learning* are still the most popular areas, however, this time, they cover "only" the 30% of the entire interest expressed by the JRC staff; the remaining 70% is well distributed among all the other areas.

Several respondents expressed interest in areas that are not her/his present working ones. Based on the expressed interest, several AI areas show an extraordinary potential growth rate, in respect to the present activity; in particular: *Motion and manipulation* (1500%), *Artificial General Intelligence* (680%), *Planning* (650%), *Robotics* (450%), *Creativity* (350%), *Social Intelligence* (322%), *Audio-visual perception* (262%).

More than *180 different keywords* were specified (most popular ones: MACHINE LEARNING, DEEP LEARNING, CLASSIFICATION, BIG DATA, TEXT MINING, and CLUSTERING). This confirms the need for a well-adopted AI ontology.

Data-driven AI at JRC

For AI activity at JRC, in the 80% of the cases the utilized data types are: *Time Series*, *Text*, *Geospatial* and *Images*. *Microblogs*, *Audio*, and *Video* data types, all together, represent "only" the 10% of the cases. The most common dataset size is *below 50GB* (57% of the cases). Only 7% of utilized datasets has a size *greater than 1PB*. These

data are commonly processed in a *Batch way* (60% of the cases); while, *Streaming Data* is utilized "only" in the 20% of the cases.

More than 75% of the JRC activities, in the AI realm, address either *Descriptive* or *Predictive problems*, utilizing *Structured* (43% of the cases) and *Semi-structured data* (33% of the cases). Most of utilized datasets (65%) come from either *Open Data repositories* (39%) or are *Originated at the JRC* (24%).

Supervised and *Semi-supervised learning* problems are the most common at JRC (42% and 23% of the cases, respectively). *Unsupervised* learning concerns the 31% of the activities and *Reinforcement learning* "only" 4%.

More than 200 pre-processing and *learning techniques* were indicated (most popular ones are: *Neural Networks*, *Clustering*, *Normalization*, and *Classification*).

Enabling technology for AI at JRC

Data Bases (both relational and not) are the most utilized storage technology, at the JRC, covering about the 65% of the AI activities. For about the 25% of the cases, *file system* is still used for data storage.

All the proposed *software technologies* for data analysis and processing are utilized at the JRC. However, only few of them have a significant *level of adoption*, which is always less than 15% (namely, *Python libraries*, *R*, *Matlab*, *Tensorflow*, and *Keras*).

In most of the cases (82%), data is *processed locally* or on *JRC clusters*. "Only" in the 6% of the cases, an *external* (high performance) *computing platforms* is utilized (i.e. Amazon AWS, Google Earth Engine, and Microsoft Azure).

AI Capacity building at JRC

The survey respondents provided references to *36 Publications* and *27 Projects* on AI. Noticeably, more than 50% of the respondents did not provide any reference.

More than 50% of the respondents did not recognize any useful service, provided by the JRC, for their work on AI. For the others, *Computing and storing Infrastructures* and *Software tools and licenses* are the most useful services provided by JRC, presently.

As for recommended additional services for supporting AI at JRC, the provision of *high performance computing infrastructures* and *training courses/seminars* represent more than 60% of the suggestions. Noticeably, more than 25% of the suggestions recommend the creation of a *JRC common facility on AI* (i.e. *AI Competence Center* / *AI Service Desk* / *AI Community of Practice*). While, the remaining 14% suggests to create a *JRC Data and Knowledge Hub*, including harmonization and tagging functionalities.

For the *JRC CoP (Community of Practice) on AI & BD*, the possible contributions concern the following activity areas: *Application & Use Case* (35%), *Data & Tools* (24%), *Seminars & Courses* (16%), *Notes Writing* (8%), *WP Chairing* (7%), *Governance* (6%), and *Legal & Policy practices* (5%).

Survey Scope

The survey scope was threefold: (1) to detect the existing expertise on AI at the JRC; (2) to understand the AI areas of interest and potential development at JRC; (c) the perceived gaps and needs to fully embrace AI at JRC.

Preliminary results of this survey were presented at the AI workshop at the JRC (AI@JRC abbreviated), which was held in Ispra on the 23rd of May 2018 –see the report: Nativi S. and Gómez Losada A. (Ed.), *Artificial Intelligence at the JRC*, 2019.

Survey and Analysis methodology

The target population was the JRC staff. The participation at this survey was voluntary, all the JRC Directorates were invited to participate. The survey instrument was an online questionnaire, published on the *EUSurvey* platform, to be filled individually.

The questionnaire consisted of 29 questions; most of them were closed multiple-choice questions with a set of answer options. Others were open multiple-choice questions, containing a blank field to introduce a specific answer that was not one of the options proposed by the questionnaire. There was no obligation to respond to any of the questions.

Collected responses were aggregated and statistical results presented, at the JRC level.

All the graphics produced from the survey responses (at JRC level) are presented in the Appendix A of this document.

Editors and Contact Information

Stefano NATIVI (Big Data Lead Scientist of JRC)
Unit B6, Via E. Fermi 749, 21027 Ispra VA (Italy); Tel. +39 0332 785975;
email: stefano.nativi@ec.europa.eu

Alvaro GOMEZ LOSADA
Unit B6, Expo building, Calle Inca Garcilaso, 3, 41092 Seville (Spain); Tel. +34 9544-88480;
email: Alvaro.GOMEZ-LOSADA@ec.europa.eu

MAIN ANALYSIS RESULTS

SECTION A: THE SURVEY RESPONDENTS

Survey Participation

The questionnaire was completed by 108 respondents (74% men and 26% women) from 29 different Units.

Administrative Category

As to the responders category, more than 45% were Contract Agents and about 40% Administrators; Supporting Staff and Intra-muros together represent less than 15% of the answers (see Figure 1).

	Answers	Ratio
Contract Agent (CA)	49	45.37%
Administrator (AD)	45	41.67%
Support Staff (AST)	10	9.26%
Intra Muros (PRE)	4	3.7%
Seconded National Expert (SNE)	0	0%
No Answer	0	0%

Figure 1. Participation at JRC level

SECTION B: QUESTIONS ON THE ARTIFICIAL INTELLIGENCE AREAS

AI Working areas

QUESTION: IN WHICH OF THE LISTED AREAS OF ARTIFICIAL INTELLIGENCE (AI) CAN BE INCLUDED YOUR ACTUAL WORK? (MULTIPLE CHOICE ANSWER ALLOWED).

Outlines

- The collected 189 choices show a great activity and interest in the indicated AI areas.
- All the proposed macro areas are covered by one or more activities, at JRC –see Figure 2.
- "Machine Learning" and "Deep Learning" areas were indicated by about the 45% of the answers –however, the first area weights the double (30%) of the latter one (15%). "Knowledge Representation", "Big Data Stewardship", and "Natural Language Processing" together cover another 30% of the answers.
- Less than five entries characterize four areas: "Creativity", "Robotics", "Planning", "Motion and manipulation".
- Seven "Other" areas, not included in the proposed list, were introduced. Four of them are: "Economic and social impact of AI", "Philosophical approach to AI, and its mirroring effect on how we think about humans", "Augmented AI", and "Standardisation and data quality". The others may be considered as sub-areas (or narrow areas) of the proposed macro areas.

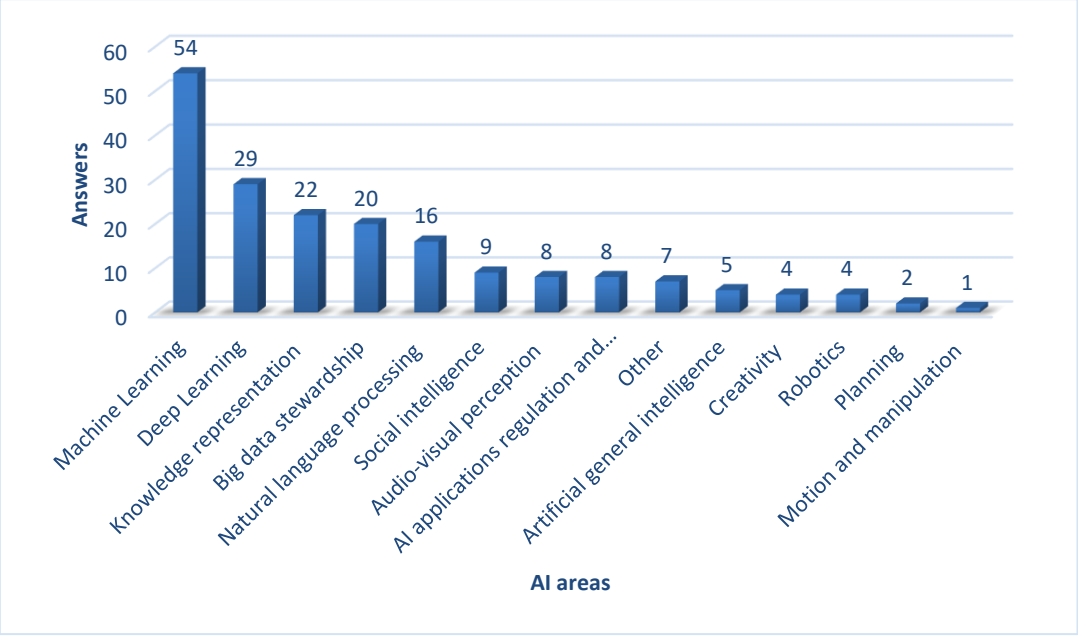


Figure 2. Present AI working areas at JRC level

AI Research Interest areas

QUESTION: IN WHICH OF THE LISTED AREAS OF ARTIFICIAL INTELLIGENCE (AI) CAN BE INCLUDED YOUR RESEARCH INTEREST? (MULTIPLE CHOICE ANSWER ALLOWED).

Outlines

- At the JRC, there exists a strong interest in AI as shown by the 439 answers collected by the survey on this question. They well cover all the proposed macro areas of interest –see Figure 3.
- "Deep Learning" and "Machine Learning" are the most popular areas, covering together the 30% of the expressed interest. The remaining 70% is well distributed among the other areas with a remarkable 8% of interest on "Artificial General Intelligence".
- Ten "Other" areas, not included in the proposed list, were introduced. They include: "Theoretical AI", "Technology law", "Artificial life", "Communicate AI", and "Expert System". The others covered applicative sector for AI –such as, security, health, forecasting, etc.

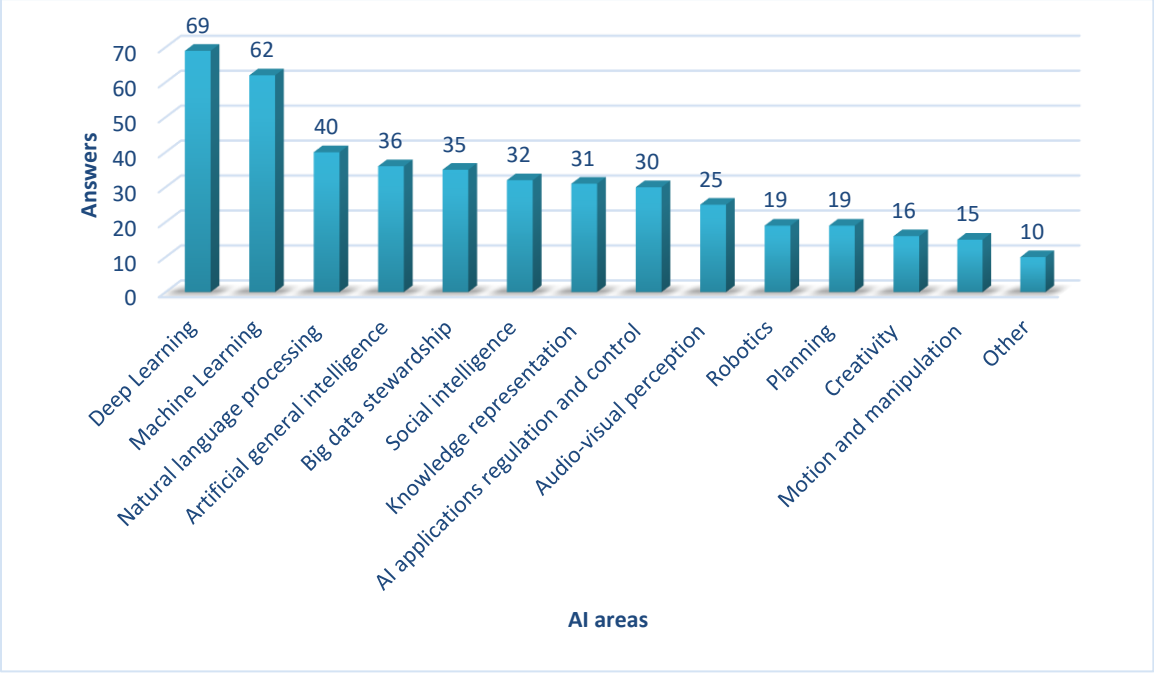


Figure 3. Research Interest on AI working areas at JRC level

AI areas: Working vs Interest analysis

OBJECTIVE: TO UNDERSTAND THOSE AI AREAS WITH A POTENTIAL ATTRACTION FOR THE JRC STAFF.

Outlines

- Several Respondents expressed interest in areas that are not her/his present working ones. This interest focused on all the proposed areas changing the popularity distribution depicted, previously –see Figure 4.
- Several areas show an extraordinary potential growth rate, in respect to the present activity, based on the expressed interest; namely: "Motion and manipulation" (1500%), "Artificial General Intelligence" (680%), "Planning" (650%), "Robotics" (450%), "Creativity" (350%), "Social Intelligence" (322%), "Audio-visual perception" (262%), "Deep Learning" (193%).

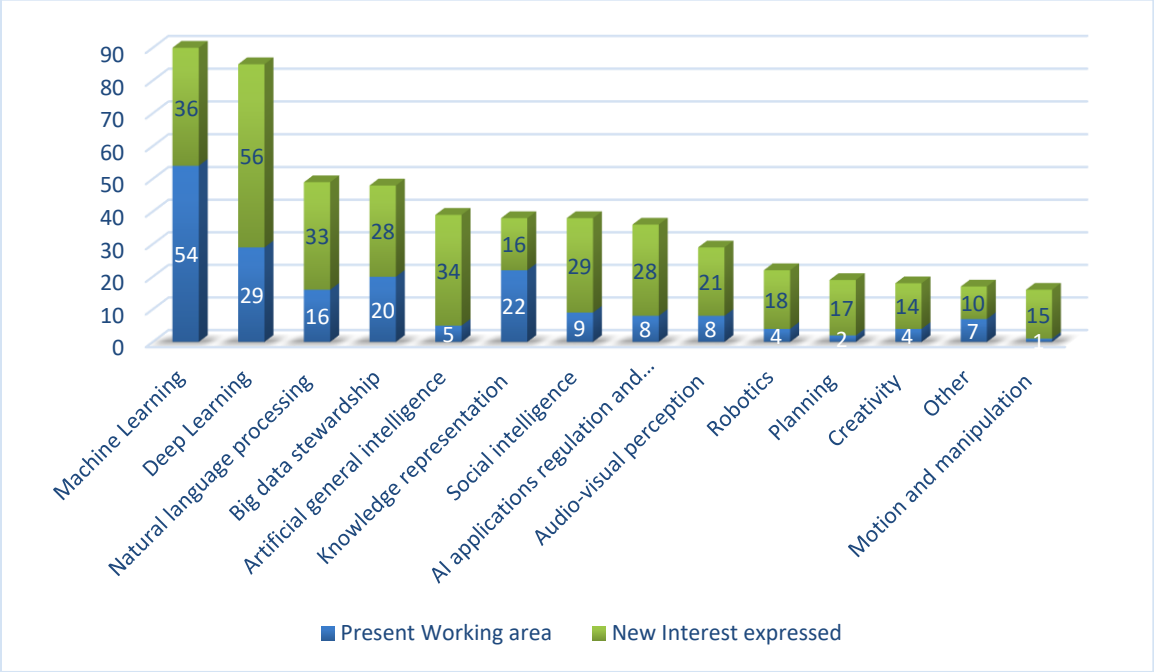


Figure 4 Potential growth of the AI areas at JRC, by comparing the present working activity and the expressed interest –by those who are not already working on the areas

Keywords

QUESTION: COULD YOU PLEASE ADD THE MOST RELEVANT KEYWORDS THAT BEST DESCRIBES YOUR WORK IN AI? (MULTIPLE KEYWORDS ALLOWED).

Outlines

- More than 180 different keywords were specified. The most popular ones are: "MACHINE LEARNING", "DEEP LEARNING", "CLASSIFICATION", "BIG DATA", "TEXT MINING", and "CLUSTERING" –see Figure 5.
- This confirms the need for a well-adopted AI ontology.

Figure 5. Most frequent keywords associated to AI at JRC

AI Common Problem types faced by JRC

QUESTION: HOW WOULD DESCRIBE THE MOST COMMON PROBLEM YOU TRY TO SOLVE WITHIN YOUR AI AREA OF EXPERTISE?

Outlines

- More than 75% of the JRC activities, in the AI realm, address either "Descriptive" (46%) or "Predictive" problems (31%). While, 13% of the activities deal with "Recommendations" –see Figure 6.
- Some 14% of the participants did not respond and another 8% selected "Other" type of activities; they include "Prescriptive" problems, "Resilience of AI algorithms", and issues dealing with "Data aggregation and harmonization".

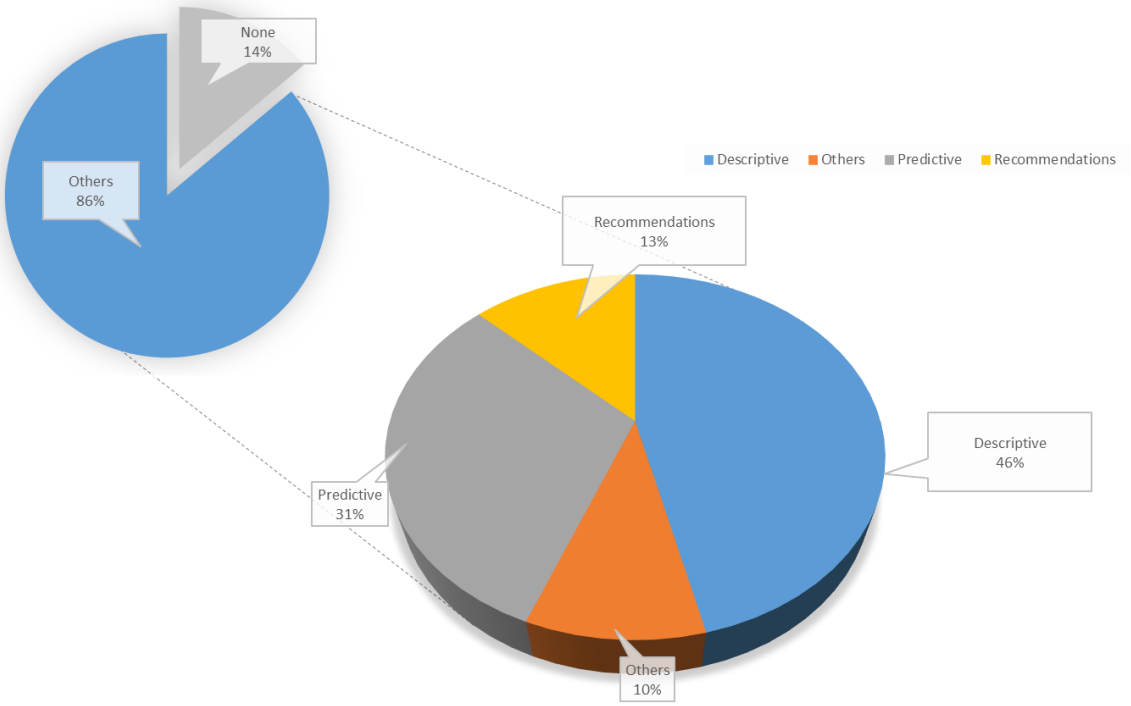


Figure 6. Types of AI problems addressed by JRC

SECTION C: QUESTIONS ON THE UTILIZED DATA-DRIVEN METHODOLOGY

GENERAL QUESTION: WHAT ARE THE CHARACTERISTICS OF THE DATA YOU USE?

Data Structure types managed by JRC

QUESTION: DATA COMPLEXITY (MULTIPLE SELECTIONS ALLOWED).

Outlines

- "Structured" data are the most utilized at the JRC (43%). Typical Web data ("Semi-structured" ones) follow being used by the 33% of the respondents. "Unstructured data" (e.g. online documents, blogs, etc.) is utilized by about 23% of the JRC activities in the AI realms. –see Figure 7.
- Some 20% of respondents did not specify any kind of data complexity.

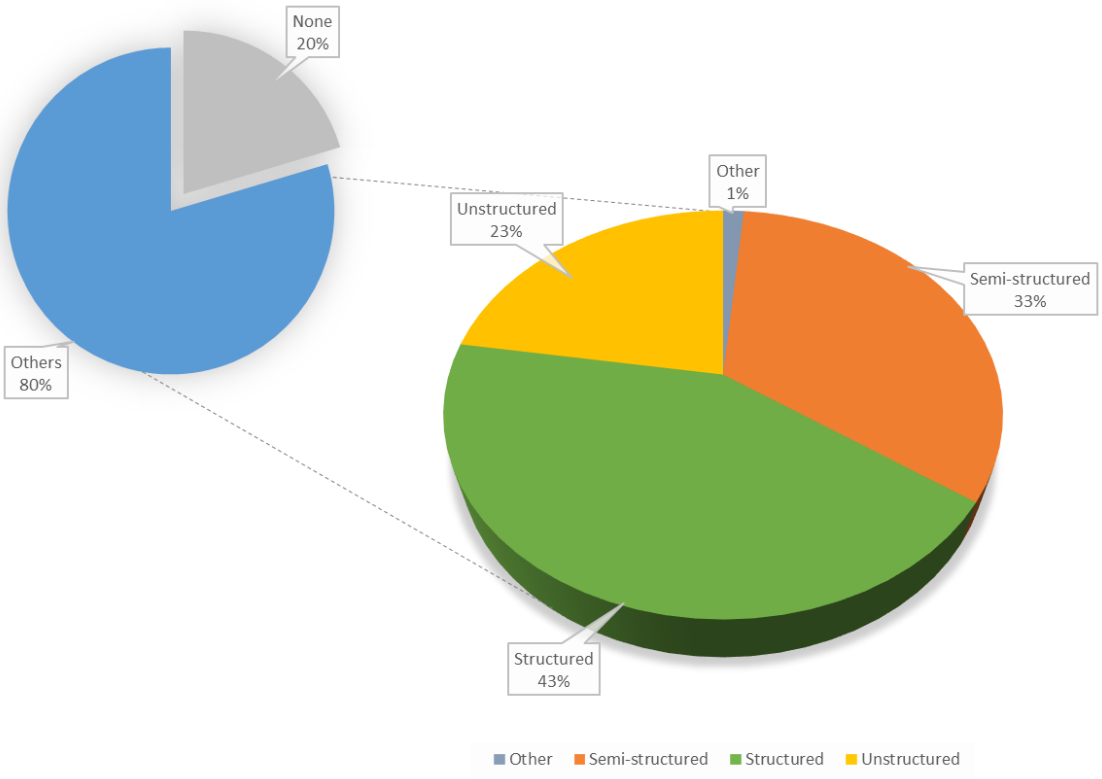


Figure 7. Complexity types of data utilized by JRC in AI applications

Data Origin

QUESTION: DATA ORIGIN (MULTIPLE SELECTIONS ALLOWED).

Outlines

- Open Data repositories are the most utilized sources by the JRC (39%). Summed to the data "Originated at the JRC" (24%), they cover about the 65% of the data used in AI activities –see Figure 8.
- About 20% of data for AI are bought by JRC/EC.
- About 15% of data is originated by "Other" sources, including: social media platforms (e.g. Facebook, LinkedIn), Web scraping, international programmes and initiatives, EU programmes and initiatives, behavioural data collected in the field of human-robot interaction.
- Some 17% of respondents did not specify any kind of data origin.

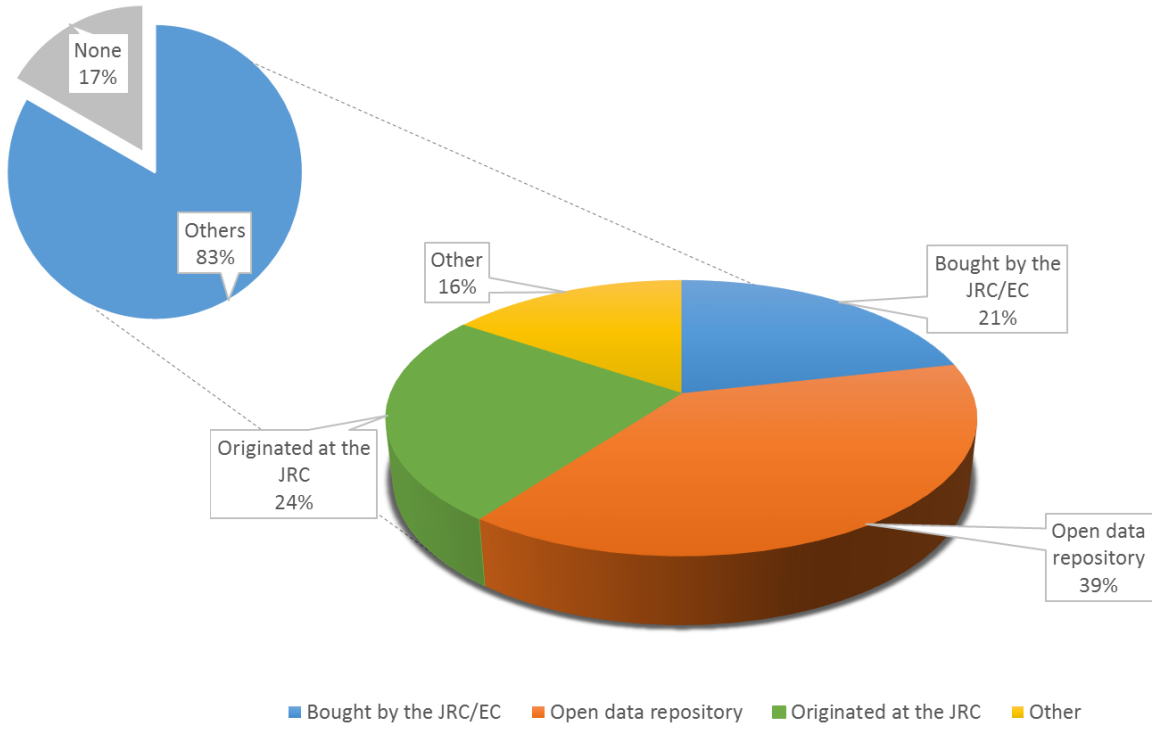


Figure 8. Origin of Data utilized by JRC in AI applications

Data Size

QUESTION: DATA SIZE (MULTIPLE SELECTIONS ALLOWED).

Outlines

- For AI activity at JRC, the most common dataset size (57%) is below 50GB, about half of them are below 50MB –see Figure 9.
- Around 35% of used datasets are greater than 50GB but less than 1PB. Only 7% of utilized datasets for AI studies has a size greater than 1PB.
- Some 20% of respondents did not specify any kind of data size.

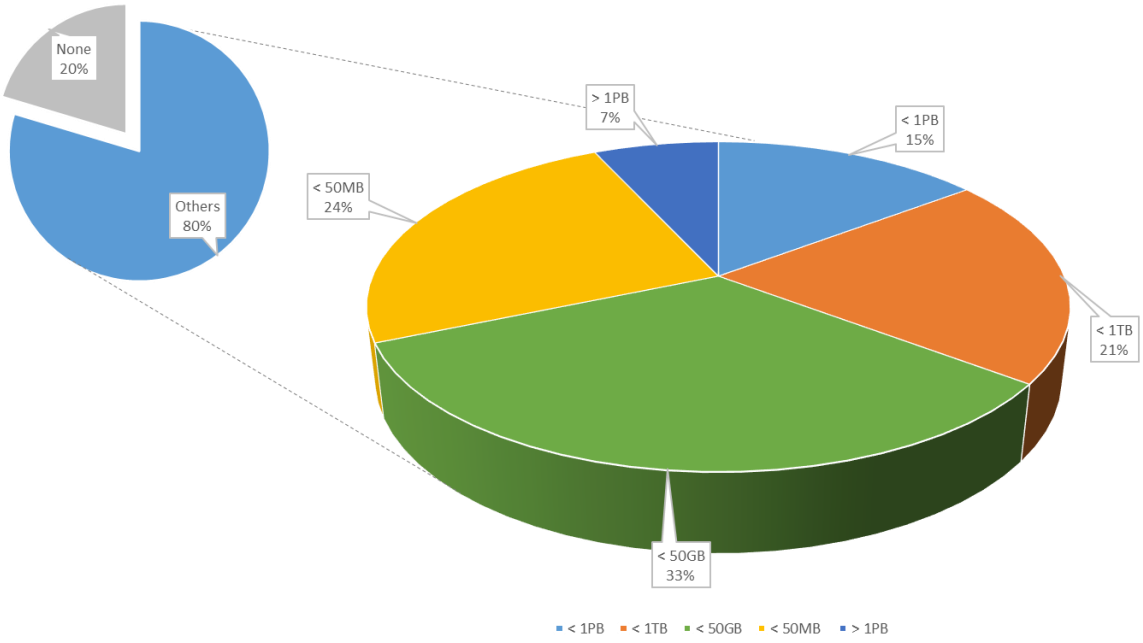


Figure 9. Size of Data utilized by JRC in AI applications

Data Processing methodology

QUESTION: DATA PROCESSING (MULTIPLE SELECTIONS ALLOWED).

Outlines

- More than half of the datasets, used at the JRC for AI, adopts a "Batch data" processing (60%); while, "Streaming Data" is utilized in the 20% of the cases. However, in 20% of the cases, the respondents did not know the kind of processing. –see Figure 10.
- Some 24% of the respondents did not answer.

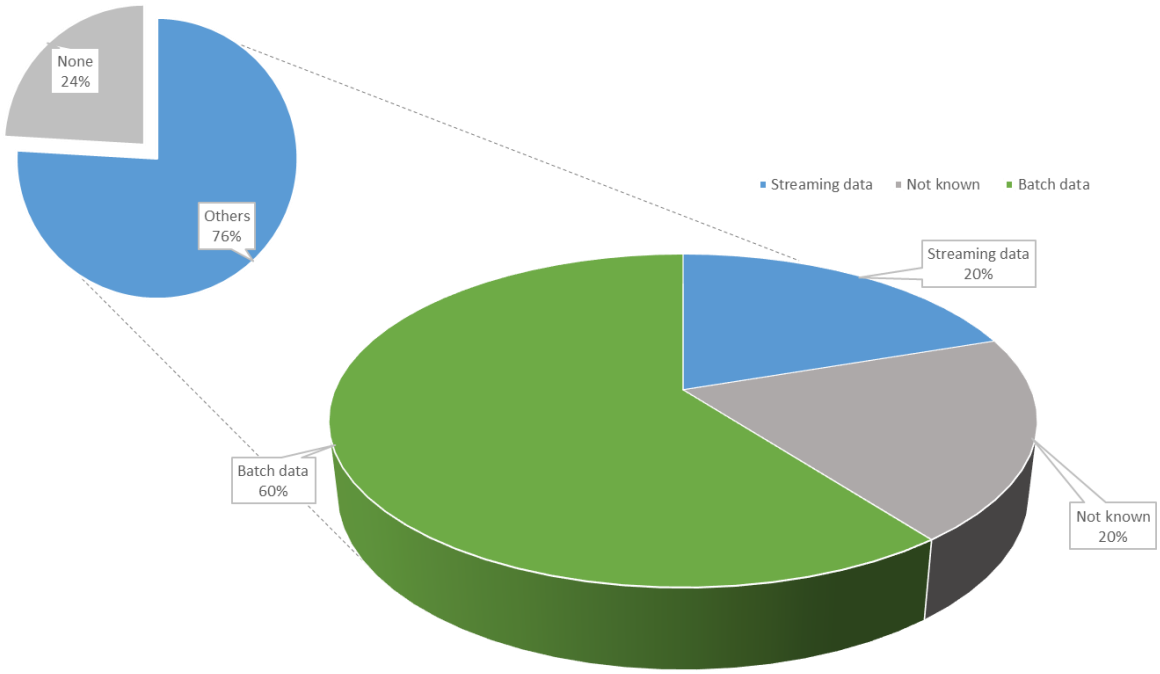


Figure 10. Data Processing methodology utilized by JRC in AI applications

Data Types/Heterogeneity

QUESTION: DATA TYPE (MULTIPLE SELECTIONS ALLOWED).

Outlines

- For AI activity at JRC, the most common data types are "Time Series" (26%), "Text" (22%), and "Geospatial" (18%) and "Images" (14%). All together represent 80% of the cases –see Figure 11.
- "Microblogs", "Audio", and "Video" data types are the least frequently used by JRC for AI –all together score less than 10%.
- As for "Other" data types, specified by the Respondents (4%), they include: "Cross-Sectional", "Structured", "Statistical", "Experimental", and "Numerical" data.
- About 20% of respondents did not specify any data type.

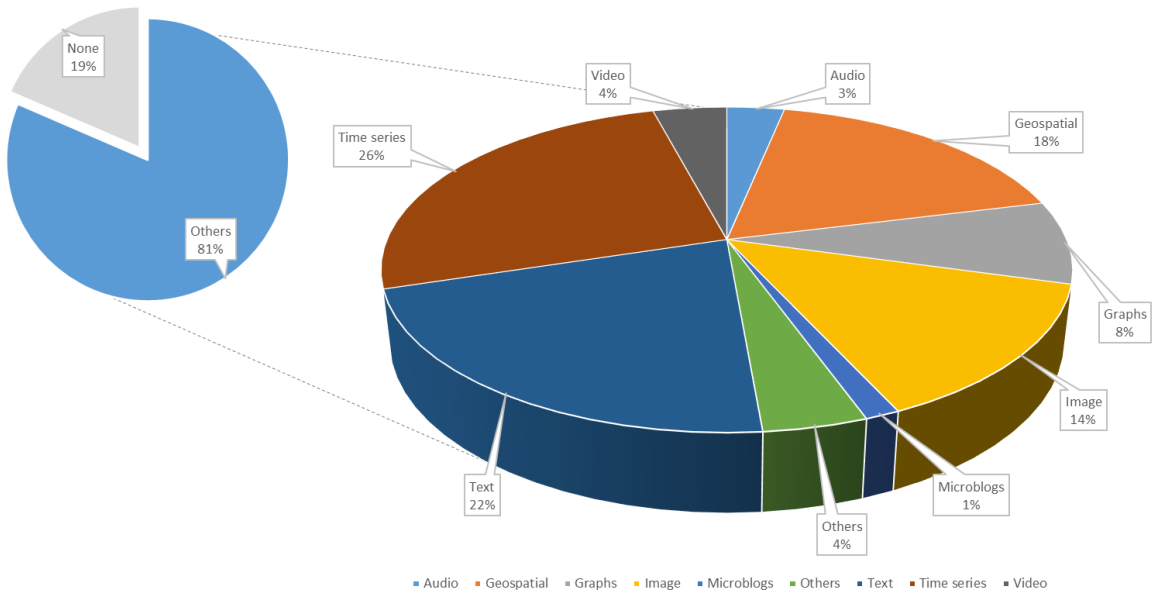


Figure 11. Heterogeneity of Data utilized by JRC in AI applications

Data Storage Types

QUESTION: STORAGE TYPE (MULTIPLE SELECTIONS ALLOWED).

Outlines

- "Relational DB" is the most utilized storage type (39% of the cases), followed by "Non-relational DB" (27%); together they represent more than 65% of the responses –see Figure 12
- For more than 10% of the cases the store technology is "Not Known"; while, in more than 20% of the cases "Other" storage technologies are used –they mostly consist of a file system storage.
- Some 24% of Respondents did not answered.

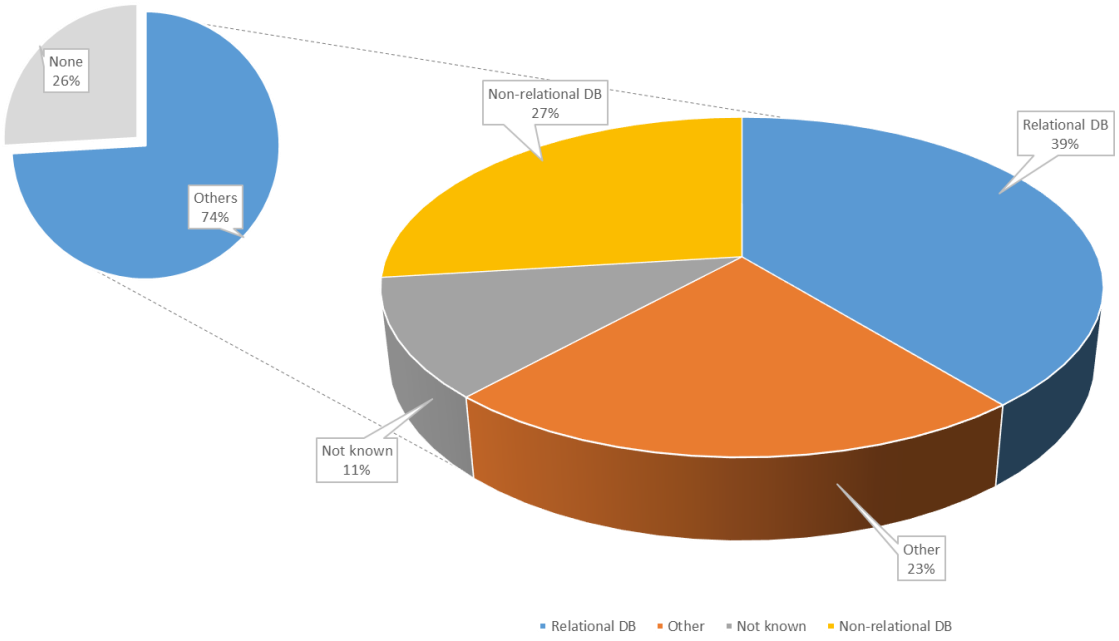


Figure 12. Type of Data Storage utilized by JRC for AI activities

Software for Data Analysis/Processing

QUESTION: WHICH SOFTWARE DO YOU USE TO PROCESS OR ANALYSE YOUR DATA? (MULTIPLE SELECTIONS ALLOWED).

Outlines

- All the proposed instruments are utilized at the JRC. However, only the following ones have a significant level of adoption: "Python libraries" (14%), "R" (13%), "Matlab" (13%), "Tensorflow" (9%), and "Keras" (7%) –see Figure 13.
- A remarkable 12% of the responses introduced "Other" instruments (more than 30), including: neo4j, Google Earth Engine, Microsoft Azure ML, software for spectral and image analysis, and coding languages along with dedicated libraries. The high diversity of specific software, used at the JRC, is especially significant.
- Some Respondents did not specify any software tool –10% of the analysed inputs.

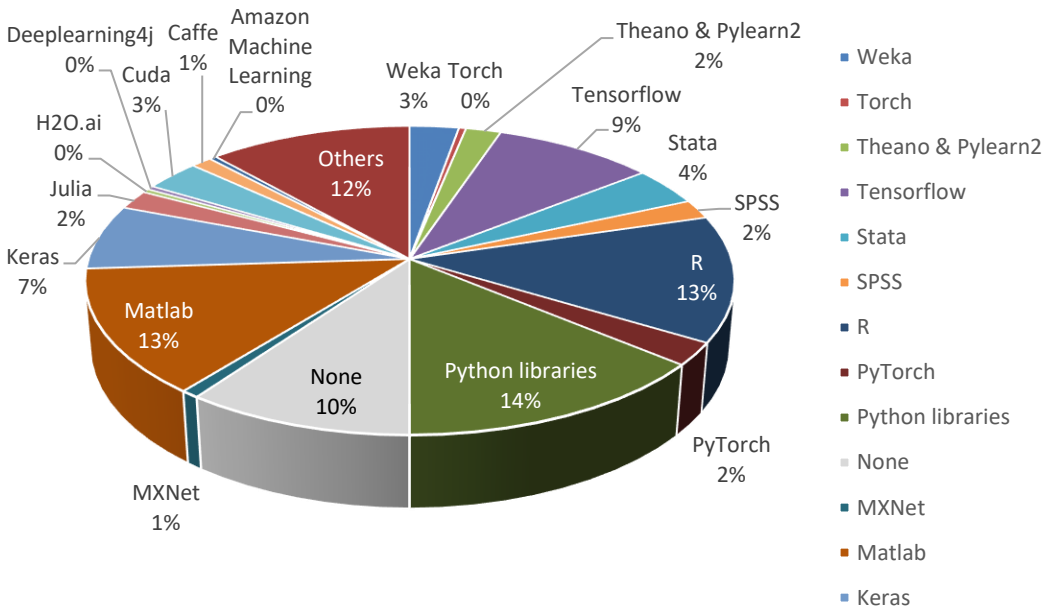


Figure 13. Software tools used to process/analyse data by JRC for AI activities

Computing Platform for Data Analysis/Processing

QUESTION: WHERE DO YOU PROCESS OR ANALYSE YOUR DATA? (MULTIPLE SELECTIONS ALLOWED).

Outlines

- In most of the cases (82%), data is processed locally (54%) or on JRC clusters (28%) –see Figure 14.
- Noticeably, 6% of the processing/analysis is done by using external computational platforms, i.e. Amazon AWS, Google Earth Engine, and Microsoft Azure.
- "Other" platforms contribute for about 12% of the cases, including "University clusters" and the "JRC Big Earth Data and Processing Platform JEODPP".
- Some 22% of the Respondents did not provide any answer.

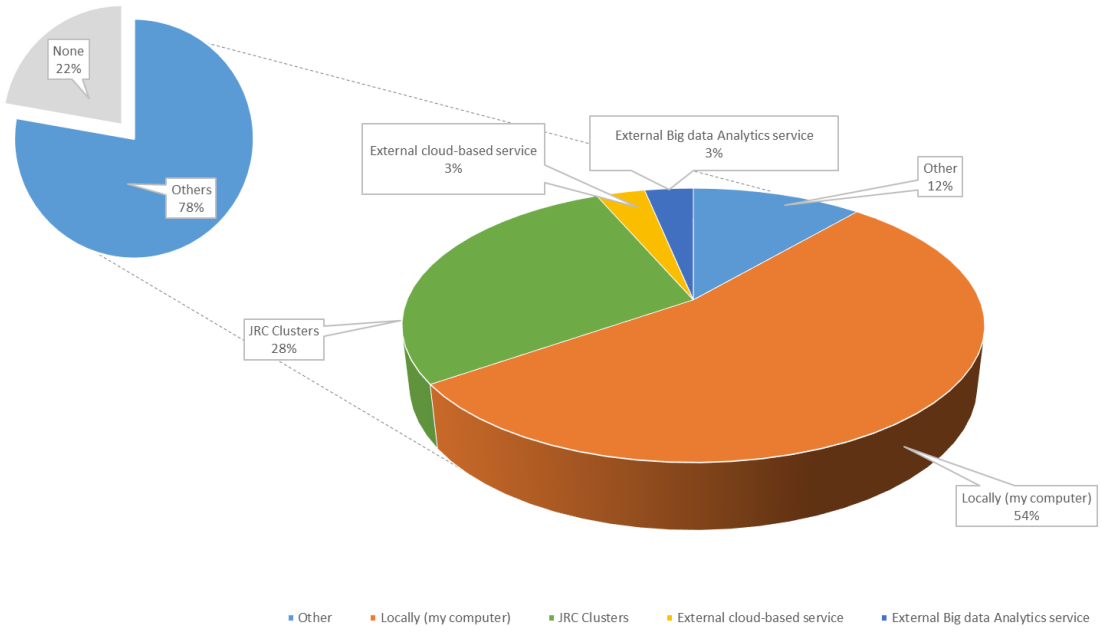


Figure 14. Computing Platform used to process/analyse data by JRC for AI activities

SECTION D: QUESTIONS ON DATA-DRIVEN LEARNING

IF YOUR WORK OR RESEARCH IS RELATED TO MACHINE LEARNING OR DEEP LEARNING

Data-driven Learning Research type

QUESTION: IF YOUR WORK OR RESEARCH IS RELATED TO MACHINE LEARNING OR DEEP LEARNING, WHERE DOES IT SUIT BEST? (MULTIPLE SELECTIONS ALLOWED).

Outlines

- "Supervised learning" problems are the most common at JRC (42%), followed by "Unsupervised learning" (31%) and semi-supervised (23%) –see Figure 15.
- Noticeably, only 4% of the reported activities deal with "Reinforcement learning".

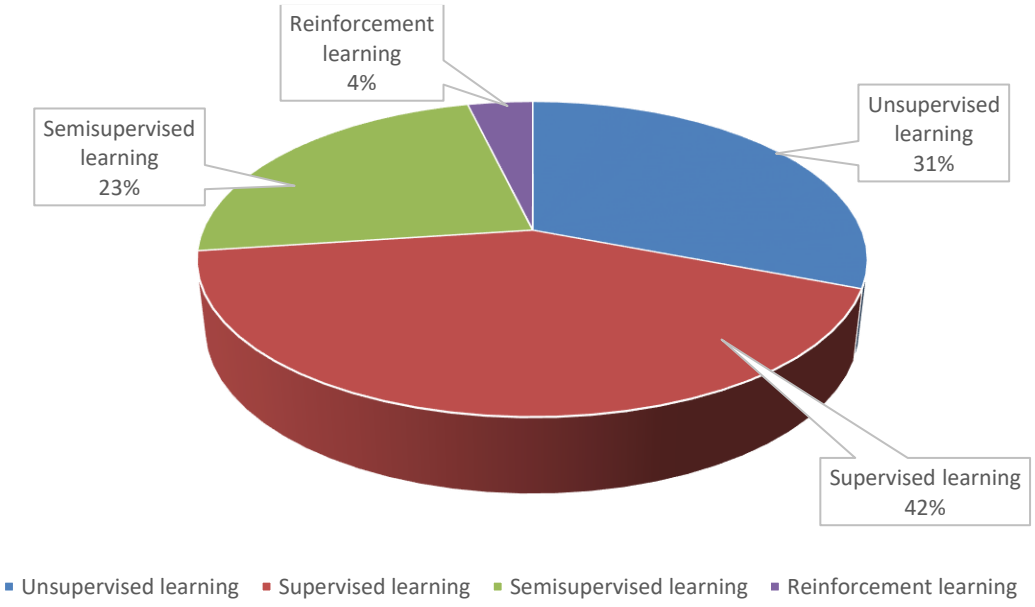


Figure 15. Type of Machine/Deep Learning activities implemented at the JRC

Keywords describing the utilized Pre-processing techniques and Learning algorithms

QUESTION: IF YOUR WORK OR RESEARCH IS RELATED TO MACHINE LEARNING OR DEEP LEARNING, WHAT KEYWORDS DEFINE BEST THE PRE-PROCESSING TECHNIQUES (IF ANY) AND LEARNING ALGORITHMS THAT YOU USE FOR ANALYSING YOUR DATA? (MULTIPLE KEYWORDS ALLOWED).

Outlines

- More than 200 pre-processing and learning techniques were inserted. "Neural Networks", "Clustering", "Normalization", and "Classification" techniques are the most popular ones –see Figure 7.

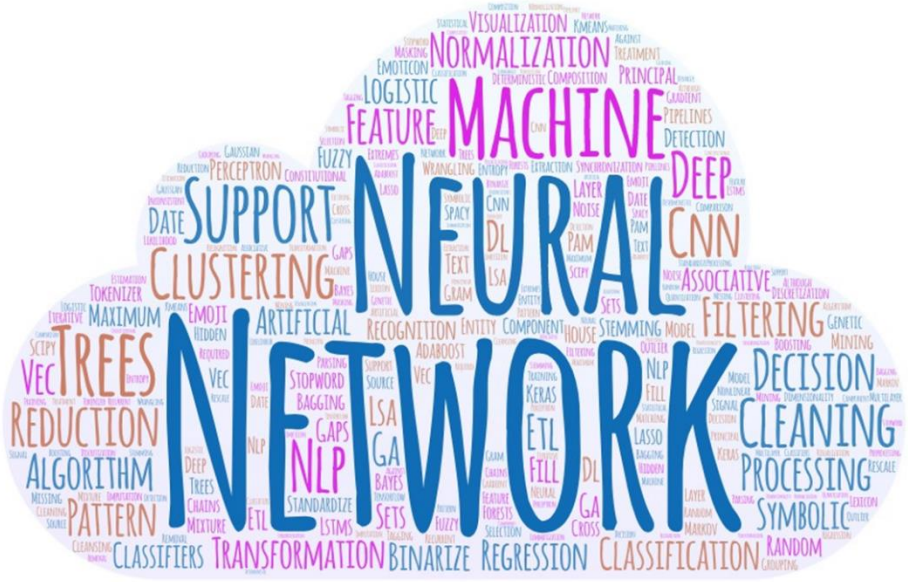


Figure 16. Type of Machine/Deep Learning pre-processing and/or learning algorithms implemented at the JRC

SECTION E: CAPABILITIES QUESTIONS

Relevant JRC Projects and Publications on AI & BD

QUESTION: COULD YOU PLEASE CITE YOUR ONE OR TWO MOST RELEVANT PROJECTS AND/OR KEY PUBLICATIONS IN RELATION TO AI? (MULTIPLE ENTRIES ALLOWED).

Outlines

- A total of 63 entries were provided: 36 are Publication references, and 27 Projects.
- Noticeably, 55 Respondents (more than 50%) did not provide any publication or project reference.

Relevant JRC Services for AI & BD activity

QUESTION: WHICH JRC SERVICES FACILITATE YOUR WORK (E.G. INFRASTRUCTURES; MODELS; SOFTWARE TOOLS; TRAINING COURSES; SEMINARS; ETC.)? (MULTIPLE ENTRIES ALLOWED).

Outlines

- More than 50% of the Respondents did not recognise any useful service for their work on AI. For the others, "Computing and storing Infrastructures" (43%) and "Software tools and licenses" (38%) are the most useful services provided by JRC –see Figure 17
- Noticeably, 13% of the Respondents indicated that "Training course and Seminars" are extremely useful for facilitating the work on AI.

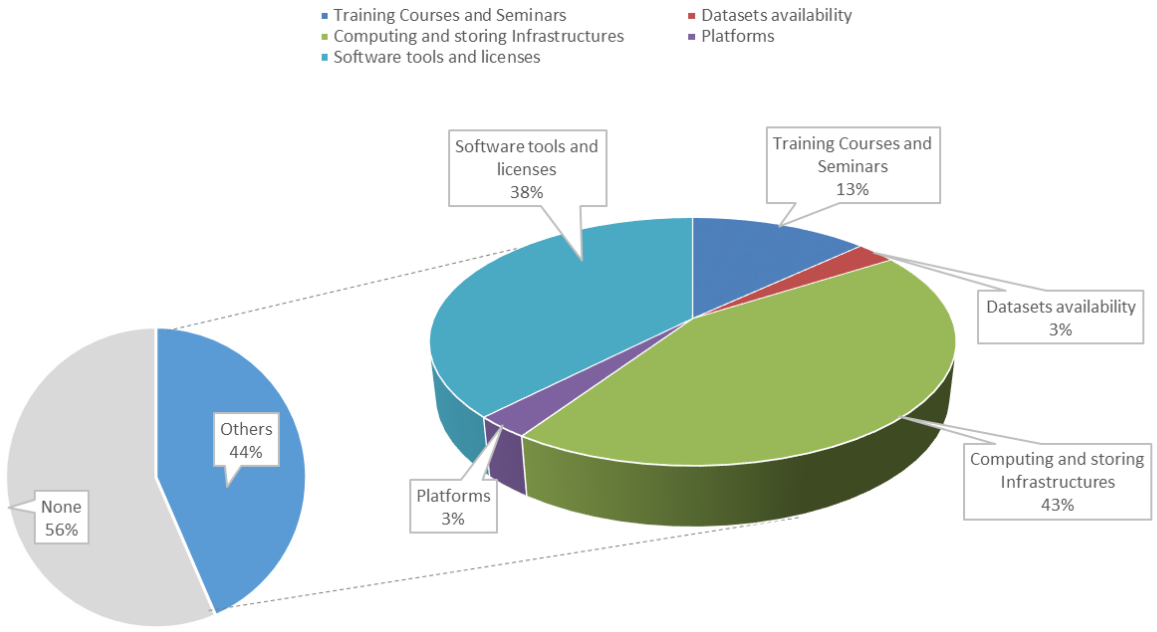


Figure 17. JRC services that facilitate the work on AI

Additional JRC Services needed for AI & BD activity

QUESTION: WHICH ADDITIONAL JRC SERVICES WOULD YOU NEED TO FURTHER FACILITATE YOUR WORK? (MULTIPLE ENTRIES ALLOWED).

Outlines

- More than 40 suggestions were collected from the 30% of the Respondents.
- The suggested additional services can be clustered, recognizing that: "high performance computing infrastructures" (35% of the responses) and "training courses/seminars" (26%) represent more than 60% of the suggestions –see Figure 18.
- Noticeably, more than 25% of the suggestions deal with the creation of a JRC common facility on AI (i.e. "AI Competence Center" / "AI Service Desk" / " AI Community of Practice"). While, the remaining 14% suggest to create a JRC "Data and Knowledge Hub", including harmonization and tagging functionalities.

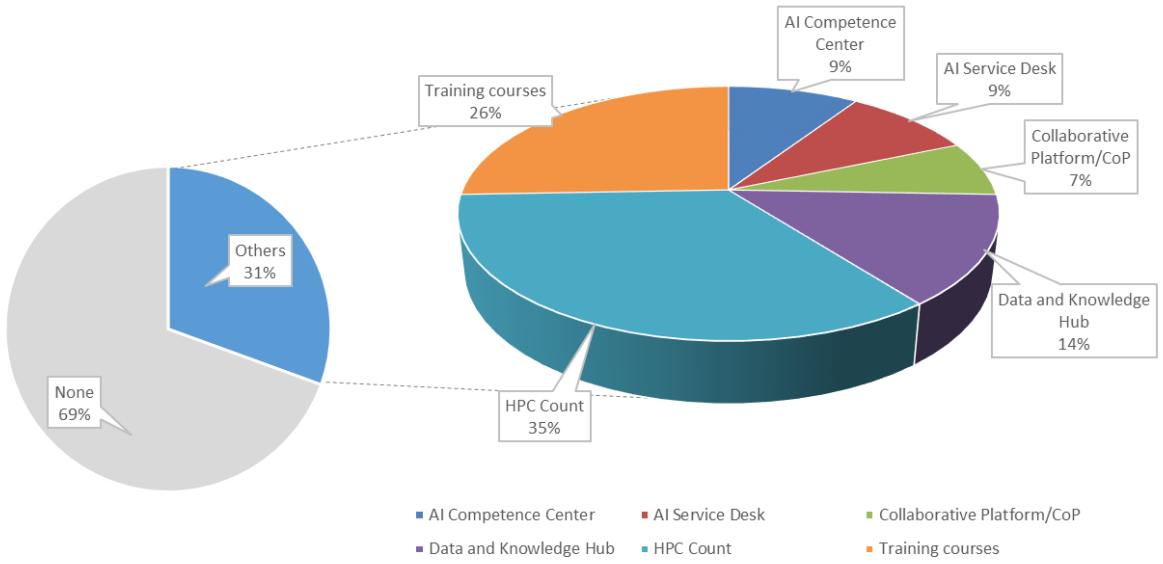


Figure 18. Additional JRC services that would facilitate the work on AI

SECTION F: QUESTION ON A COMMUNITY OF PRACTICE ON AI AND BD

Possible contributions to the AI & BD Community of Practice

QUESTION: WHAT ARE YOUR POSSIBLE CONTRIBUTION(S) TO THE AI & BD COMMUNITY OF PRACTICE (E.G. CHAIRING A WP; SHARING SOFTWARE OR MODEL; PROVIDING SEMINARS; WRITING SOME NOTES; REPORTING A SUCCESSFUL PILOT OR USE CASE; ETC.)? (MULTIPLE ENTRIES ALLOWED).

Outlines

- More than 100 suggestions were provided by the 42% of the Respondents.
- The possible contributions to the CoP, can be clustered, in the following activity areas (percentage of the received suggestions is reported): "Application & Use Case" provision (35% of the suggestions), "Data & Tools" provision (24% of the suggestions), "seminars & Courses" (16% of the suggestions), "Notes Writing" (8% of the suggestions), "WP Chairing" (7% of the suggestions), "Governance" (6% of the suggestions), and "Legal & Policy practices" (5% of the suggestions). –see Figure 19.

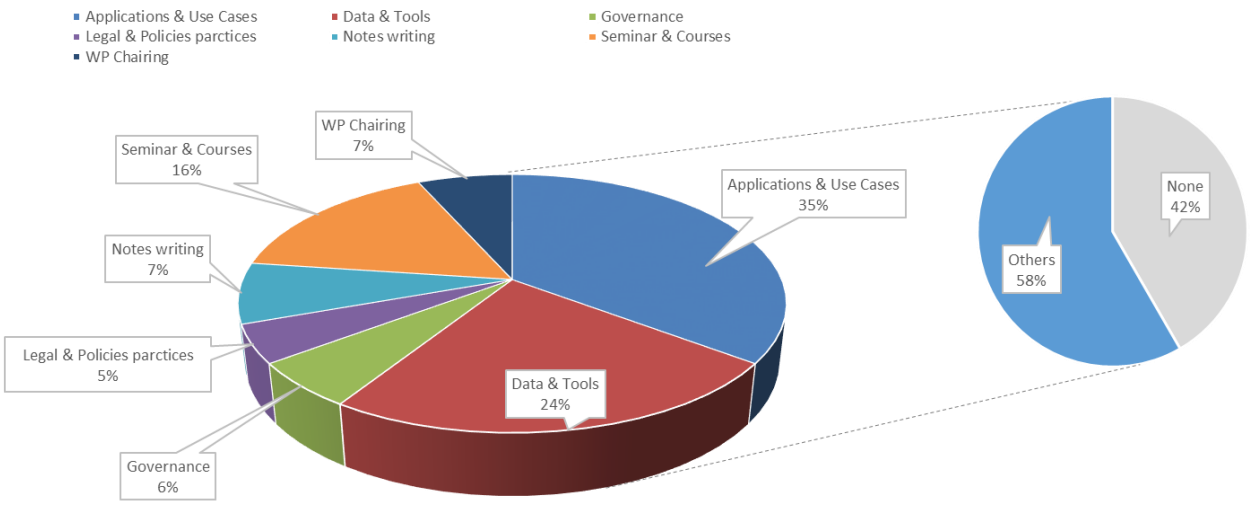


Figure 19. Possible contributions to the JRC CoP on AI & BD.

Acknowledgement

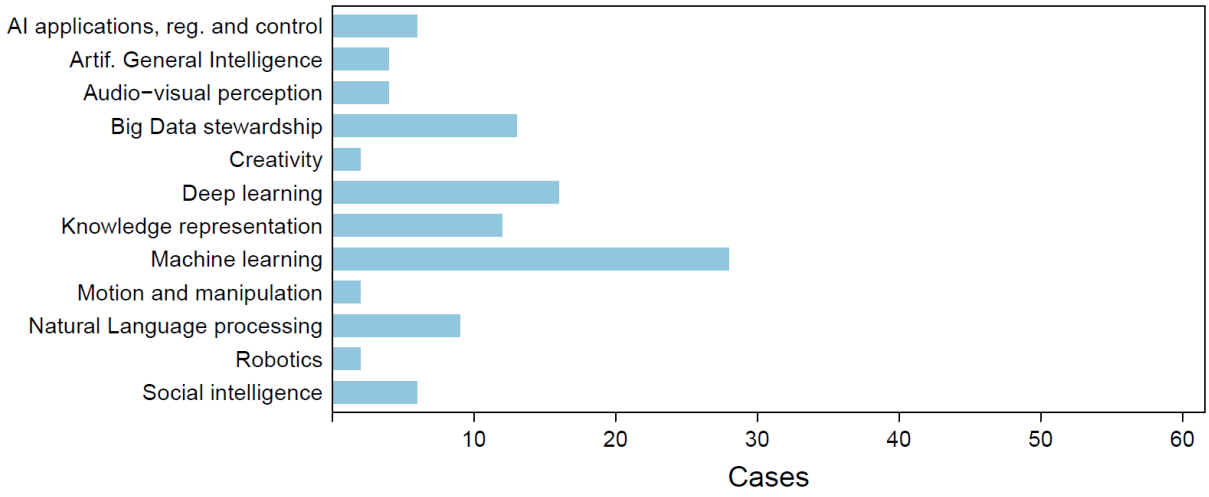
We are particularly grateful to all the participants to the survey connected with the first workshop on Artificial Intelligence at JRC.

We also want to thank Karen Fullerton and Loizos Bailas for the important help in the survey preparation and management as well as all the other JRC colleagues that made good suggestions for this successful event.

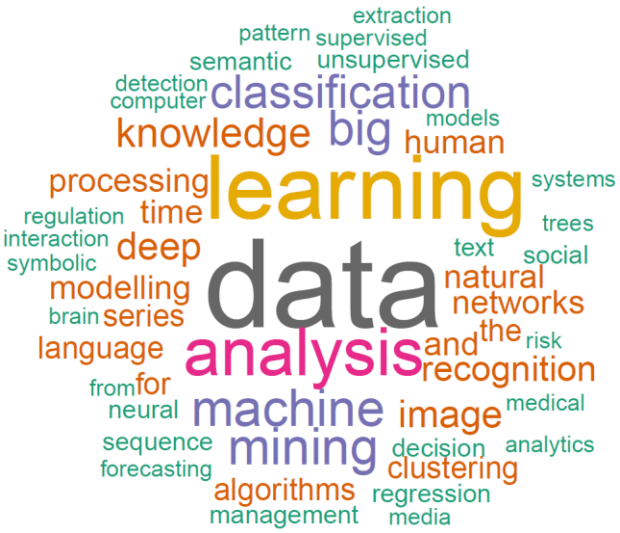
Annex: Questionnaire Entries Count

Questions 1-4. In which of the next areas of Artificial Intelligence (AI) can be included your actual work or your research interest?

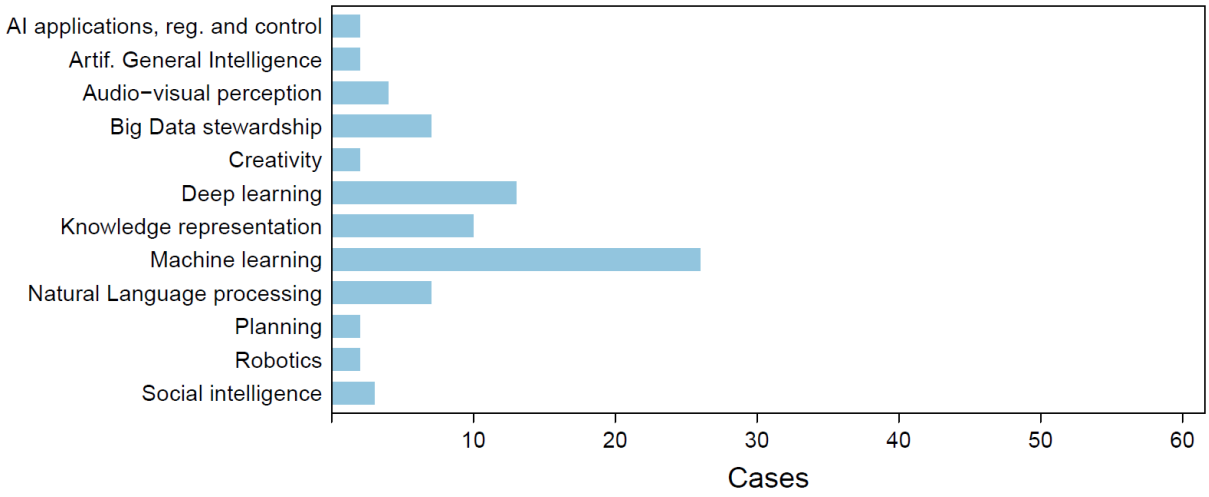
Areas of work in AI



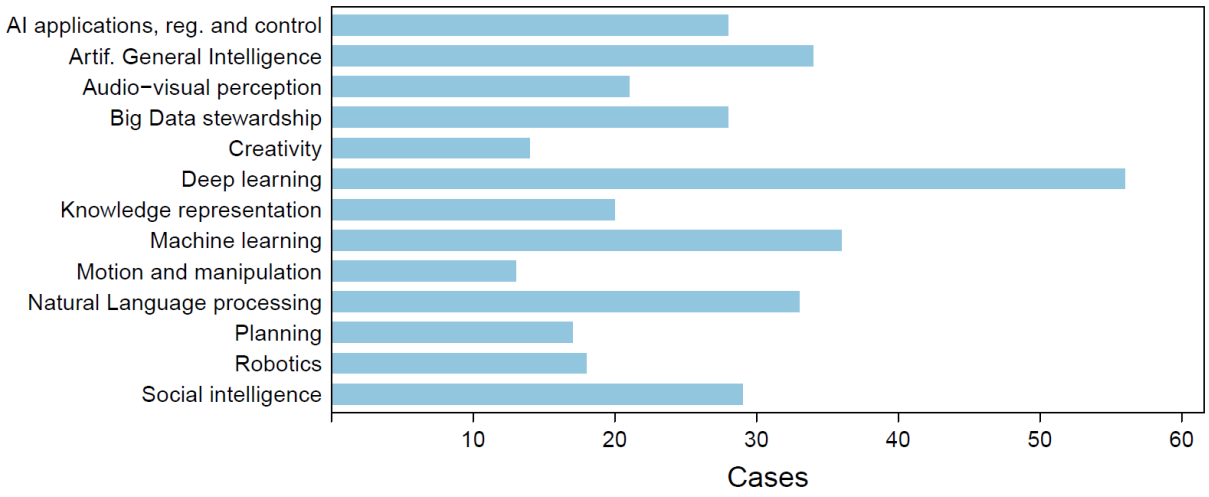
Other area(s) not listed that I am working on.



Areas of work and interest in AI

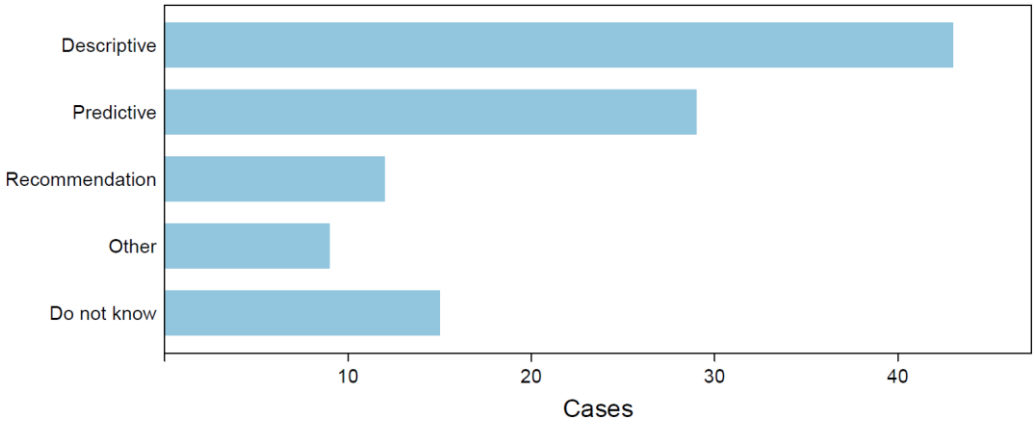


Areas of interest in AI



Other area(s) not listed that I am interested

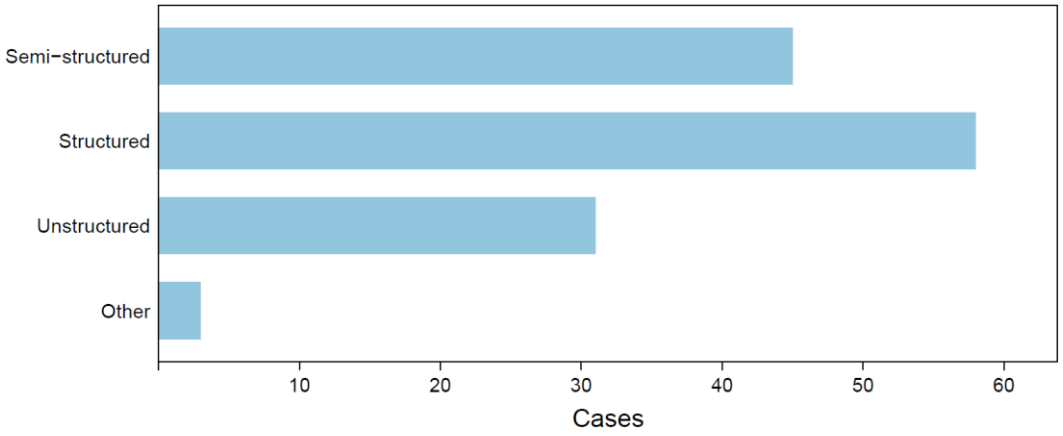
Question 5. How would you describe the most common problem you try to solve within your AI area of expertise?



Question 6. Other common problems you try to solve within your AI area of expertise different from Descriptive/Predictive/Recommendations.

- Data cleaning, normalization, data semantics
- Data heterogeneity
- Descriptive and predictive
- Resilience of AI algorithms against malicious actions, transparency of algorithms (explainability), application of AI in cybersecurity, privacy and data protection in AI
- Combination of "descriptive" and "recommendations"
- Industrial Modernisation, Agri-Food and Energy
- Decision support (helping nuclear inspectors)
- Descriptive, predictive, prescriptive

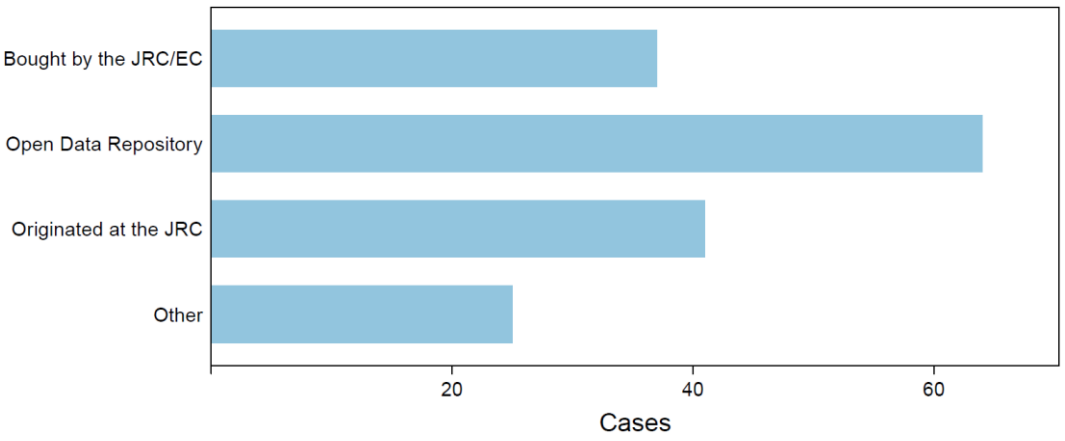
Question 7. If following a data-driven methodology, what are the characteristics of the data you use? (you may need to tick more than one box at each category)



Question 8. Other characteristics of data different from Structured/Semi-structured/Unstructured categories.

- A mix between structured data (from sensors e.g.) and unstructured data (from inspectors' "perception")

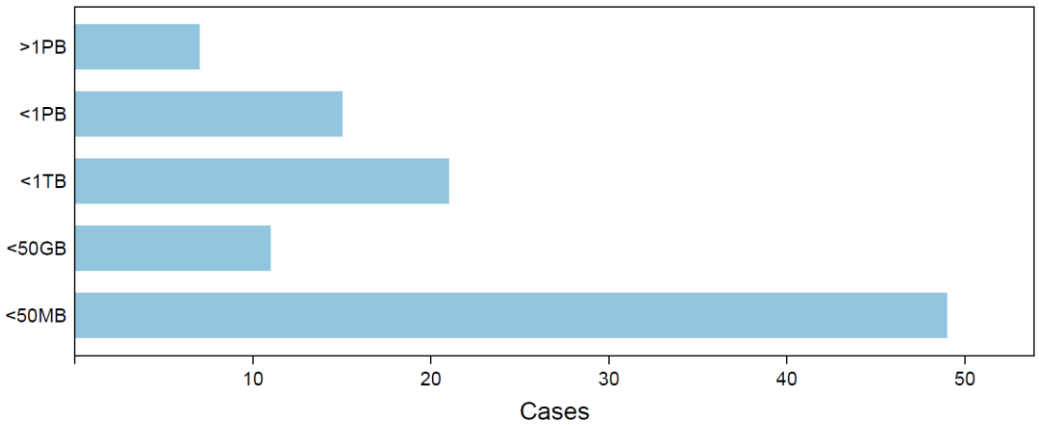
Question 9. Origin of data.



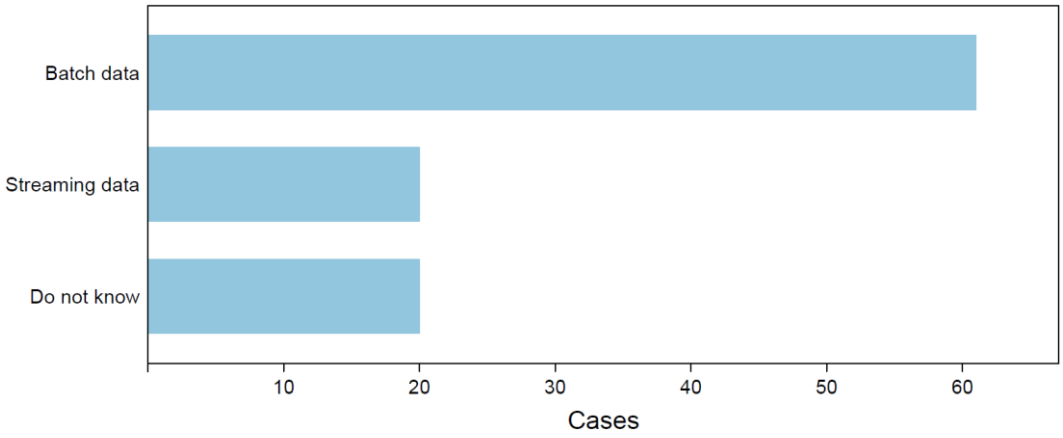
Question 10. Other origin from Open data repository/ Bought by the JRC/ Originated at the JRC.

- Visual documents (photos, films) being found in the above mentioned 8 to 13 km of JRC documents
- Behavioural data collected in the field of human-robot interaction (A5)
- IoT and sensor data
- Collected from sensors, inspectors and possibly open data (e.g., from media monitoring)

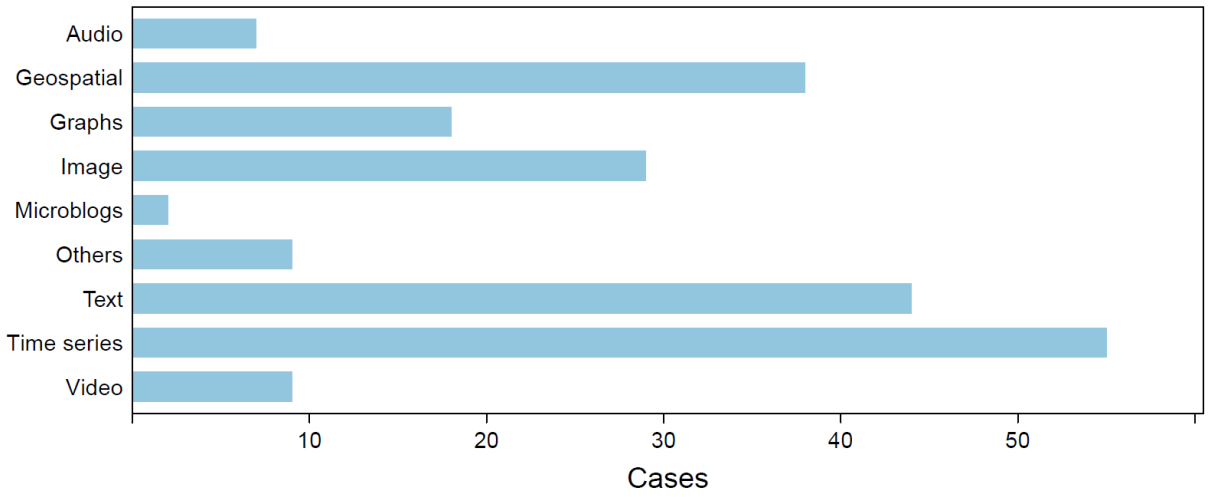
Question 11. Size of data.



Question 12. Processing.



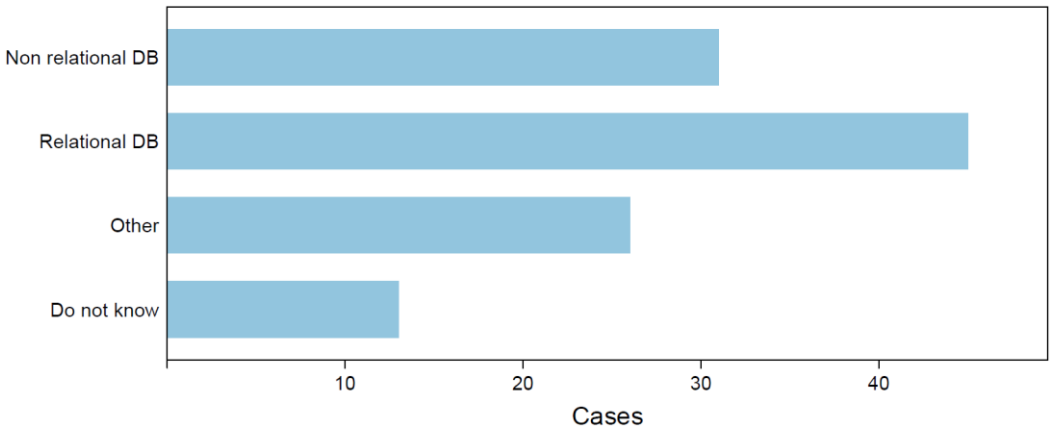
Question 13. Type of data.



Question 14. Other type of data different from Audio/Image/Video/Geospatial.

- Cross-sectional data.

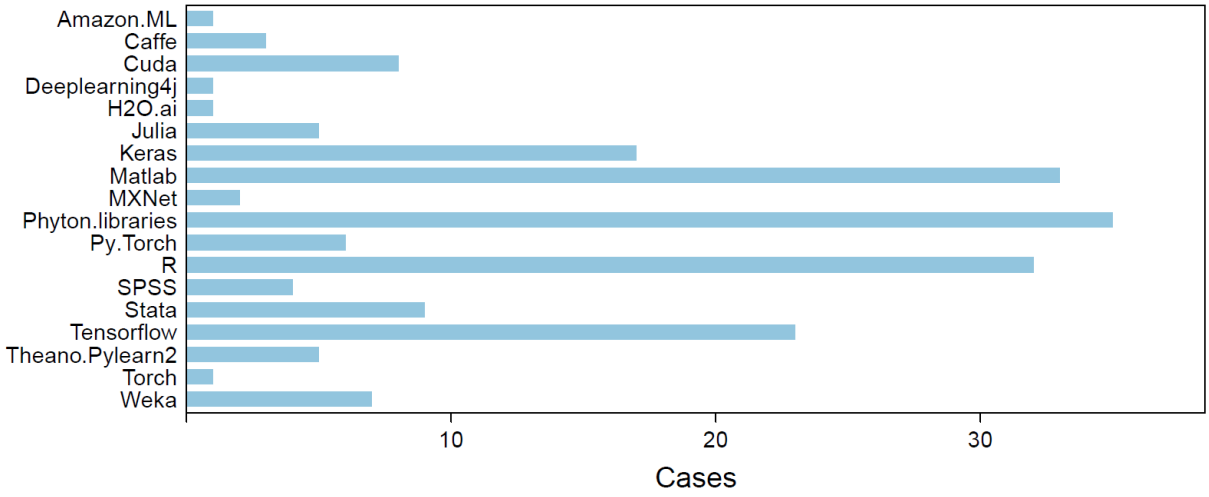
Question 15. Storage type.



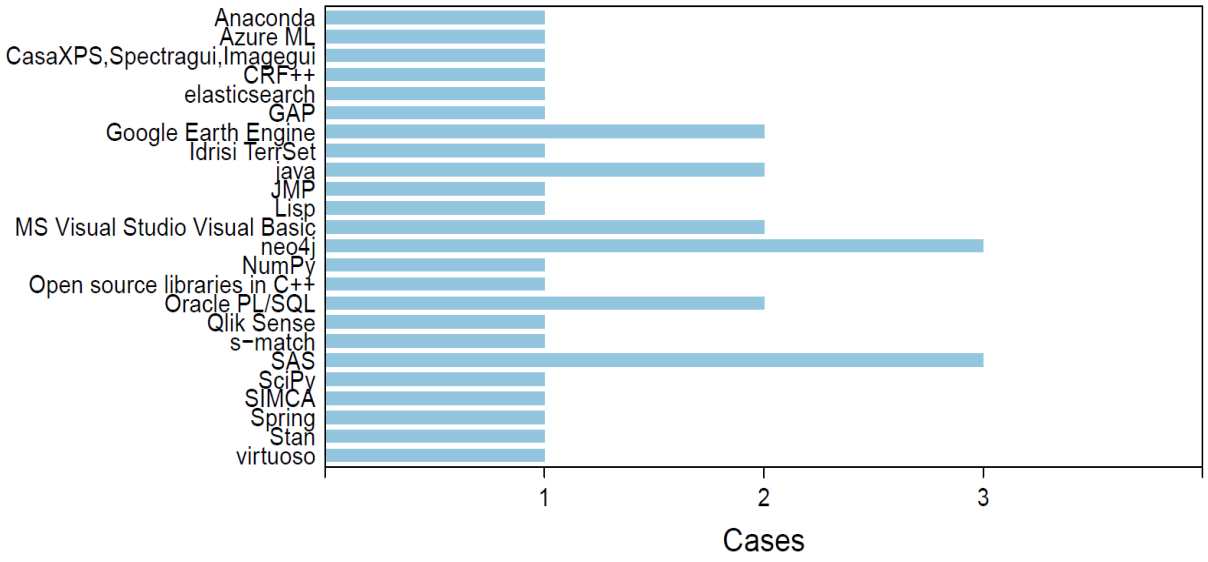
Question 16. Other type of storage type different from Relational DB/Non-relational DB/Other/Not known.

- Mostly file hierarchies; Flat files; File system; NetCDF; Text files
- Image formats; Image file formats; Geospatial file formats
- Flat files referred to in database
- Graph database
- Web-based distributed
- Cloud storage, disk storage
- Data mart
- Spatial layers

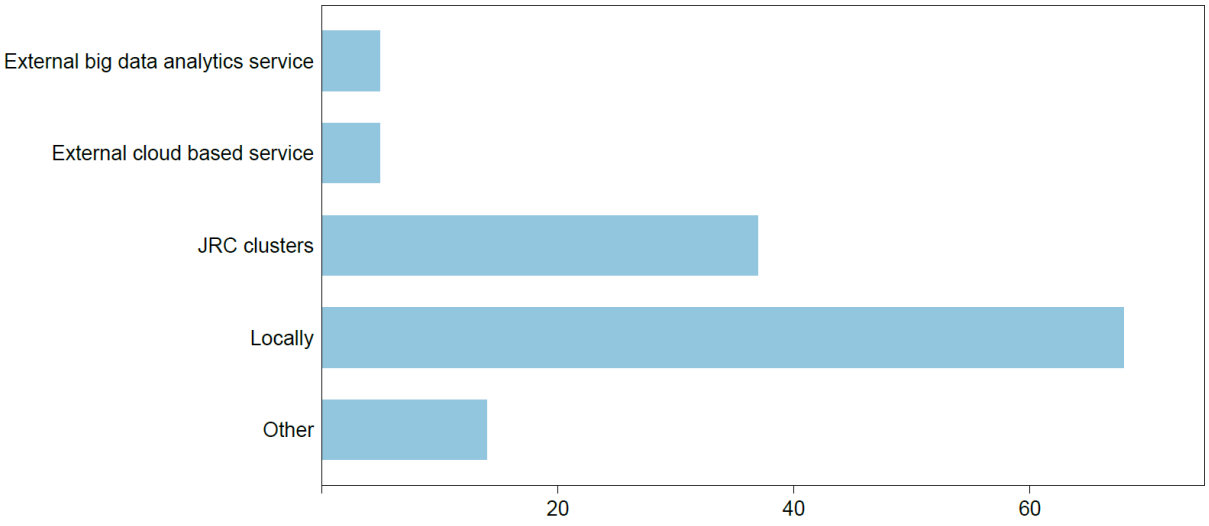
Question 17. Which software do you use to process or analyse your data?



Question 18. If other (software). Please, specify



Question 19. Where do you process or analyse your data?



Question 20. Where do you process your data different from Locally, JRC Clusters, External cloud based service (which one?), etc?

- Amazon web Services (AWS)
- Amazon EC2

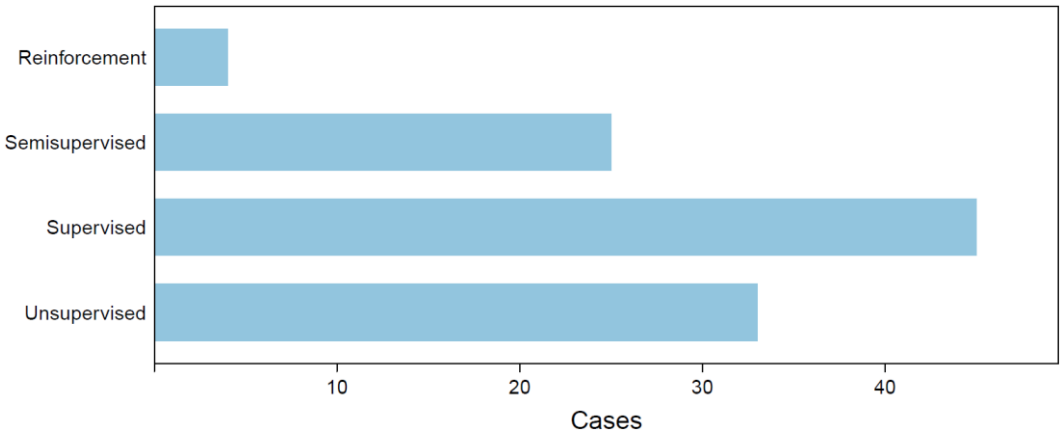
Question 21. Which Big Data analytic service?

- Google Earth Engine
- Azure ML

Question 22. Which other Big Data analytic service?

- Virtual server; Remote server; University storage
- Local servers; Lab servers and clusters; Common drive; Oracle database server
- JRC Big Earth Data and Processing Platform JEODPP

Question 23. If your work or research is related to Machine Learning or Deep Learning, where does it suit best?



Question 24. Other kind of learning different from Supervised learning, semi-supervised learning, unsupervised learning, reinforcement learning, other.

- No answers provided.

Question 25. If your work or research is related to Machine learning or deep learning, what keywords define best the pre-processing techniques (if any) and learning algorithms that you use for analysing your data?

Pre-processing techniques

- Data cleaning
- Data/text normalization
- Data Imputation
- Data binarization
- Data integration
- Data transformation
- Data reduction, clustering, quantization, grouping, principal component analysis
- Deterministic algorithms
- Data resampling
- Data standardization
- Data masking
- Outlier detection
- Data visualization
- Data cross-comparison (from multiple sources)

- Data wrangling (to overcome data originators mistakes)
- Long data processing pipelines
- Pre-Deep Learning pattern recognition and application back to data to improve quality
- Natural Language Processing pre-processing algorithms
- Data composition
- Data synchronization
- Data randomization
- Data discretization
- Dimensionality Reduction and Compression Algorithms

Learning algorithms:

- Genetic Algorithm, GA-Kmeans, PAM for extremes, nonlinear support vector
- Classification algorithms and regression trees (CART), Random Forest, Bayesian Networks, multilinear regression, k-Nearest neighbours, hierarchical clustering, principal component analysis
- Sequencing, recurrent neural networks
- Associative analysis, symbolic machine learning
- Symbolic machine learning, robust against noise and inconsistent training data, no pre-processing required, associative analysis learning
- Deep constitutional networks, Support vector machines
- Supervised sequential pattern learning, unsupervised sequential learning, statistical assessment of knowledge
- KERAS, SPACY, SCIPY, Word2Vec
- Topic modelling fitting and learning: variational inference algorithm, Gibbs sampling
- Emoticon/emoji treatment, stemming/lemmatization, special lexicon matching, feature extraction, feature computation, feature selection, word embeddings, LSA, Support vector machines, Logistic regression, Maximum entropy, Multilayer Perceptron, Decision Trees, Naïve Bayes, Meta-classifiers, AdaBoost, Bagging, Deep Neural Networks
- Recognition, Tagging, Parsing, tokenizer, clustering, classification
- Symbolic Machine Learning, In-house developed classifiers, Tensorflow
- Lasso, random forests, artificial neural networks
- Tokenization, word2vec, Convolutional neural networks, Natural language processing, stopwords
- Gradient boosting, artificial neural networks
- SVM, artificial neural network
- Statistical signal processing, Gaussian Mixture Model
- Deep learning, CNNs, LSTMs, Markov Chains, Support vector machines
- Classification
- Regression techniques, artificial neural network, classification
- Maximum likelihood estimation, principal components
- Fuzzy data sets, multi-layer perceptron
- Variational Neural Networks, Generative Adversarial Networks, LTSM Recurrent Networks, PCFGs, Hidden Markov Models, Posterior Sampling, Variational Bayesian Inference

- Decision tree, clustering, neural network

Question 27. Which JRC services facilitate your work (e.g. infrastructures; models; software tools; training courses; seminars; etc.)?

- Training courses on forecasting with machine learning
- Infrastructures, especially HPC and GPUs
- Access to a cluster where to run models, with large amounts of memory and Linux
- NVIDIA GPU clusters
- AI Competence Centre -as a complementary corporate service to the Commission
- Cloud services
- Training; access to more suited infrastructures
- Training courses; seminars
- European Media Monitoring
- Logistics to operate the initial digitization. iHiP text analysis to establish data sets for training and evaluation of ML generated algorithms
- Courses on advance statistical analysis of data with unknown quality
- Central JRC/EC knowledge repository with JSON-producing RESTful API for data enrichment
- Training, improved computer power –hardware
- Training courses and seminars, software tools and infrastructure for deep learning applications or access to e.g. Amazon Cloud or Google Cloud Machine Learning -although data privacy can be an issue when using such services
- Computer science support in advanced computing solutions -in this domain-, hardware specific acceleration expertise -e.g. GPU- + adequate training
- Have an institutional account for large scale processing on Google earth Engine platform especially for exploiting the Google Cloud Storage
- JRC should consider establishing an AI "service desk" for its scientists, so that the scientists don't have to go into the nitty-gritty details of implementing AI themselves, if they don't want to
- Training courses on AI
- An application that enables live work sharing, such as overleaf or google docs
- Computing cluster –GPU-, python libraries and training, Deep learning seminars, resources to create dataset
- Elastic Computing + Big enough Data Infrastructure
- GPU cluster; sharing of datasets, techniques and know-how
- A more consistent way to access and use JRC data - which are currently spread across different systems, and are not represented in a harmonised way
- I would need a server with a GPU, which I currently don't have
- Collaboration with JRC scientists working on AI and using AI. Help to apply AI tools to my work
- More training in machine learning
- Technical workshops
- Data processing (number crunching), but not storage
- Data from European Media Monitoring (EMM) and competences in sentiment analysis

- Tie up with blockchain, kick off of a fail fast xlab to prototype use of such technologies on ourselves
- A network (e.g.: on connected) for discussing similar experiences would be very useful
- Training, improved computer power –hardware
- Any JRC service that could provide support to the Interregional Partnership dealing with Artificial Intelligence and Human Machine Interface -AI&HMI
- JRC entities (competence centers?) specialised in multi-source data aggregation (using AI?) and in setting up blockchain infrastructure

Question 28. What are possible contribution(s) to the AI&BD Community of Practice (e.g., chairing a WP; sharing software or model; providing seminars; writing dome notes; reporting a successful pilot or use case; etc.)?

- Reporting a successful pilot or use case, sharing models, writing some notes and participation at CoP
- I offer my availability to contribute to project exploring the societal dimension of AI
- I can contribute to any of the possibilities: chairing a WP, sharing software or model, providing seminars, writing some notes, reporting a successful pilot or use case, etc.
- Chairing a WP, providing seminars & writing some notes
- Chairing a WP on AI in the medical field, writing some notes
- I am not sure. I would need to see the structure and functioning of the AI & BD Community of Practice first
- I intend to give a course next year related to Deep learning for sequential data
- Legal: IP advice, software licensing expertise. Project coordination/reporting in Brussels if required
- Sharing software or model, providing seminars, contributing to experimental work on benchmarking AI methods in support to EO data classification
- Symbolic machine learning algorithm
- Forensic investigation over digital media and source camera identification/verification/content retrieval. Software development for automatic image detection/recognition
- Write notes, report a successful pilot or use case
- Providing seminars, consultancy, reporting of successful use cases -if privacy clauses permit
- Standardisation, semantics, data access and quality
- Reporting use cases, conducting seminars
- Everything from the following: sharing software or model, providing seminars, reporting a successful pilot or use case, managing an AI & BD use case, providing consultation
- Chairing a WP, reporting successful case studies
- Writing notes, presentation, reporting on pilot study
- Providing seminars, reporting use cases
- Providing training and evaluation data sets, i.e. digitized document sets that are manually processed in a way that needs to be replicated by the algorithm in order to allow a semi-automatic classification of documents and their storage in a searchable database
- Reporting successful or unsuccessful pilot and use cases

- Sharing software
- Consuming use cases, providing NLP-related services
- Case studies presentations, conduct seminars
- Sharing software or models, writing notes, reporting successful use case, chairing WP
- Participation, sharing
- AI/BD is part of a set of technologies for practical use. That's the main driver. Please resolve mandate overlap with JEODPP. Willing to contribute to "community of practice"
- Sharing software and providing trainings on our in-house developed tools for machine learning
- Conducting Human-robot interaction behavioural studies, organizing workshops, Conference Program Committee chair, Editorial work, Publishing in relevant conferences and journals
- Providing seminars, reporting research in the HUMAINT project
- Providing a seminar, sharing some software or models, supervising some research, reporting a successful pilot or use case
- Report on the JRC Big Data platform (JEODPP)
- Sharing knowledge and software where/if possible
- sharing software, reporting use case
- helping and advising on the legal or intellectual property aspects of a project or of a field of research
- Sharing software and models, reporting use cases
- Sharing expertise, data and tools
- Input to research communities on AI
- Writing articles, reproducible research: open source code and data
- Collaboration in pilot case in the field of nanotechnology, material science
- Sharing software or model, writing some notes, reporting a successful pilot or use case
- Providing seminars, chairing a WP
- Sharing software or model
- My interests are in machine learning, not Big Data. Any possible contribution should become more clear at the end of the workshop
- Applications for education, education policies
- I would be please to contribute to the setting-up of new AI project using text mining and expanding my knowledge in R and Python to be able to provide support to users in AI field
- I find it difficult to see the relevance of AI in my work. Perhaps it could help in reading the model equations, that's all I can see
- Reporting a successful case and sharing with other potential help in the JRC staff to publish in an international journal of AI or other
- Providing and organizing seminars (much needed!), sharing software, writing notes or lectures
- Sharing our results from the ongoing activity on heterogeneous sensor networks and text and data mining
- Sharing Results and math methods in analysis of complex systems
- I guess I could organise some bootcamps. Happy to help where I can... we need in particular a better framework bringing together AI, blockchain and STS/analog (real world) governance and power structures to work out what is possible and what we really want as a society.

- Poor in this moment, seen that we have good intentions (and concrete needs) more than experience. Already bringing our use case, that may be common with other departments, can be a little contribution.
- Mapping of competences and matching of business opportunities in the area of AI&HMI, reinforce European Value Chains in this field, demo projects, successful pilot cases etc
- Reporting on successful pilots, chairing a WP, writing notes
- Sharing experience from large-scale initiative of MS data sharing and sharing expertise in knowledge management and modelling
- Use case on application of AI (and blockchain) to nuclear safeguards (and non-proliferation)
- I would be interested in scientific ML, where not only correlation but the cause-effect is investigated. Furthermore finding and characterizing rare events from time series or streaming data. I can share scenarios, applications/projects
- Out of the box thinking; help in communication to non-experts
- Sharing our work on complex networks

GETTING IN TOUCH WITH THE EU

In person

All over the European Union there are hundreds of Europe Direct information centres. You can find the address of the centre nearest you at: https://europa.eu/european-union/contact_en

On the phone or by email

Europe Direct is a service that answers your questions about the European Union. You can contact this service:

- by freephone: 00 800 6 7 8 9 10 11 (certain operators may charge for these calls),

- at the following standard number: +32 22999696, or

- by electronic mail via: https://europa.eu/european-union/contact_en

FINDING INFORMATION ABOUT THE EU

Online

Information about the European Union in all the official languages of the EU is available on the Europa website at: https://europa.eu/european-union/index_en

EU publications
You can download or order free and priced EU publications from EU Bookshop at: https://publications.europa.eu/en/publications. Multiple copies of free publications may be obtained by contacting Europe Direct or your local information centre (see https://europa.eu/european-union/contact_en).

KJ-NA-29803-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub

ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office

doi:10.2760/54605

ISBN 978-92-76-08847-9